

# 修士学位論文

論文題目 既訪問スポットと検索結果の対応付けに

基づく観光地検索システムの説明性向上

手法

氏  
ふ  
り  
が  
な  
名

潘  
はん  
健太  
けんた

専攻

情報学専攻

指導教授

北山 大輔 深教授

修了年月(西暦)

2020年3月

工学院大学大学院

# 要旨

近年、ユーザは観光旅行をする前に Web 上の情報やガイドブックを活用して計画を立てることが多くなっている。旅行計画を立てるためには、まず訪問したいエリアを決定する。その後、観光スポット検索サイトやガイドブックの観光スポット情報を調べて観光スポットを決定する。その際、目的とする観光スポットに関する概要やアクセスするための方法などの情報を集める..

Web 上の情報に関して、観光スポットに関する総合的な情報が掲載された公式サイト、過去にそこを訪れた観光者による旅行記ブログサイト、そして、評価が記載されたレビューサイトなどのさまざまなサイトがある。しかしながら、種々のサイトが混在しており、膨大な情報から自分の嗜好に合う観光スポットの情報を得ることは簡単ではない。目的の情報を収集するためには手動で閲覧し取捨選択しなければならない。ユーザの意識決定の材料の 1 つとして、観光スポット検索サイトがある。これらのサイトには特定の観光スポットを訪問したことのあるユーザがレビューを投稿し、観光スポットに関する豊富な情報が存在している。観光スポット検索サイトに掲載された情報を基に、利用者（ユーザ）は自身の嗜好や目的に合致した特徴を持つ対象観光スポットを検索する。多くの観光スポット検索サイトでは、対象観光スポットに付与されるメタデータを利用した以下のような検索方法が提供されている。

- ユーザが指定したタグを持つ対象観光スポット
- 評価点などの数値メタデータに基づく並び替え
- 入力単語（クエリ）を概要文中に含むオブジェクトの検索

しかし、これらのメタデータを用いた検索方法だけでは検索が困難な場合も存在する。たとえば、対象観光スポットを実際に利用することでユーザが得られる印象・感情・体験などを表す特徴に基づき対象観光スポットを検索したい場合などが該当する。また、検索できても訪問したいエリア内には数多くの観光スポットが検索結果として表示される。ユーザは検索エリアに関する事前知識がないため、検索された多数の観光スポットからどの観光スポットの説明やレビューを読むべきか効率的に判断することは困難で、自身の要求に合う

観光スポットを見つけることは容易ではない。

ガイドブックは、Web 上の情報に比べて整理されており、有名な観光スポットや流行のものに焦点を当て、そのスポットの料金や見どころ、そして食べ物などが掲載されている。また、ガイドブックも近年は電子化されており、技術的には検索可能である。ある観光スポットについて、複数のガイドブックが取り上げることもあり、行き方の比較検討をするためにそれらを読むのは労力を必要とする。そのため、Web 上の情報と同じ問題が起り得る。

そこで、本研究では、観光地検索システムの検索結果である観光スポットを効率的に理解するために、ユーザが訪問したことがある観光スポットを使って説明性を向上する手法を開発する。本手法を開発するにあたり課題は 2 つ存在する。

## 1. 観光スポット間の対応付け

まず、第 1 の課題として、ユーザが訪問したことがある観光スポットと、検索結果スポットをどのように対応付けると説明として効果的なのか明らかではない。ある観光スポットに対応する観光スポットを判定する手法として、レビューやカテゴリが類似したコンテンツであるという観点で判定する方法が一般的である。カテゴリなどのメタデータが類似する対応付けの場合には、大まかな利用目的レベルでの説明が可能であると考えられる。レビューが類似するという場合には、より細かい体験のレベルでの説明が可能となると考える。このように対応付けの方法によって、その説明性も変化すると考えられる。本研究では、前述の 2 つに、さらに、相対的な特徴による類似という、より詳細なレベルでの対応付けを加え、比較することで、説明に適した対応付け方法を明らかにする。

## 2. 地図上における観光スポットの対応関係の可視化

第 2 の課題は、数多くのユーザの過去の訪問先と検索結果をどのように提示すると、説明として適切なのか明らかではないことである。まず、地理的情報の可視化と意味的な関係を同時に可視化することを考える。地理情報の可視化としては、抽出した情報を地図上にマッピングするのが一般的である。また、意味的な関係性の可視化として、グラフモデル上でオブジェクト間のつながりを関連度で表現することが多い。観光スポット間の意味的な対応関係を可視化するために、検索結果スポットは地図上で実在する座標に固定し、既訪問スポットを関連度に基づいて適切に配置することが妥当であると考える。そこで、ある検索結果スポットに対して、入力した既訪問スポットがどのような関係にあるのか地図上で、一目で把握できるように、既訪問スポットと検索結果スポットの対応関係を可視化する手法を提案し、一般的な表形式や表示情報を増加させた形式と比較する。このことにより、説明として適切な可視化方法を明

らかにする。

各課題についての取り組みの詳細を説明する。1. の観光スポットの間の対応付けについては以下である。ユーザが今まで体験したことがあり、かつ気に入った観光スポットを利用し、訪問したいエリア内の観光スポットに対して関連付けすることで、手軽に観光情報を理解する説明性向上手法を提案する。まずははじめに、ユーザは体験履歴の中で今まで訪れたことがある、かつ気に入った観光スポットを入力する。次に、これから行ってみたい都道府県やエリアを入力する。これらの入力に対し、観光スポット検索サイトの観光スポットのレビューデータを使ってユーザの体験履歴と検索結果を対応付ける。すなわち、ユーザの未知な観光スポットに対する理解を支援するため、すでに訪問したことがある観光スポットの特徴を用いて、未訪問エリアの観光スポットを説明するために対応付けを行う。この考え方には、ユーザが以前に経験した物事を現在の物事に適用する一種の類推であるといえる。既知の知識（ベースと呼ぶ）から概念（ターゲットと呼ぶ）を獲得するときに類推思考が働くとされる。

具体的な対応付け手法について説明する。ある検索結果スポットの特徴ベクトルに最も類似度が高い既訪問スポットを対応付ける。類似度計算には、コサイン尺度を用いる。観光スポットの対応関係を示すだけでは、どのような点で対応するのかを理解するのは難しい。そこで、検索結果スポットと既訪問スポットの関係性を表すキーワードをユーザに提示する。キーワード抽出手順は、TF-IDF法を使って対象となる既訪問スポットと検索結果スポットの特徴語とTF-IDF値を求める。2つのスポットに共通する特徴語のTF-IDF値の調和平均を用いて、対応付けしたスポットの説明可能なキーワードとして抽出する。

本研究における観光スポットの対応付けに関して、ユーザによって、その強調すべき特徴が変化することがあり得る。たとえば、日本の有名な観光スポットを履歴に持つユーザにとっての「東京都庁舎展望台」と「金閣寺」について考える。このとき、このユーザにとっての「金閣寺」の特徴は、お寺、金色、京都などである。また、「東京都庁舎展望台」の特徴は、展望、夜景、建物、新宿などが考えられる。一方、京都の寺院を履歴に持つユーザにとっての「金閣寺」と「清水寺」について考える。このとき、同じ「金閣寺」でもお寺や京都などは特徴にはならず、このユーザにとっての「金閣寺」の特徴は、金色、金箔、輝きなどとなる。また、「清水寺」の特徴は、舞台や一望などである。このような特徴を相対的特徴と定義し、通常用いられる特徴ベクトルでの対応付けに加え、本研究では、この効果も明らかにする。

観光スポット間の対応付けを評価するために、対応関係を表す特徴語抽出評価、観光スポットの対応付けに関する評価、検索結果スポット説明の有効性評価の3つの評価実験を

行った。対応関係を表す特徴語抽出評価について、提案手法である調和平均を用いて特徴語を抽出する手法は、算術平均法と乗算法より観光スポット間の隠れた関係を表現することが可能かつ、無関係な特徴語を抽出しにくいことが確認できた。観光スポットの対応付けに関する評価について、相対的特徴ベクトルを用いて対応付けするのは、メタデータ法と分散表現法より詳細な情報をユーザに提示することができることを明らかにした。しかし、相対的特徴ベクトルでは、ユーザに提示する関連するスポットについての意外性はあるが、無関係に見える対応付けをする問題点があることがわかった。検索結果スポット説明の有効性評価について、提案手法では検索結果スポット名、対応する既訪問スポット名と対応関係を説明するための特徴語を提示しているに対し、検索結果スポット名のみとの比較を行った結果、提案手法の方がユーザにとって観光スポット間の関連性が連想しやすく、旅行を計画するのに有用であることがわかった。

2. の地図上における観光スポットの対応関係の可視化については以下である。地図上で既訪問スポットと検索結果スポットの対応関係と特徴語を可視化することで、ユーザが検索結果スポットに対する理解を容易化することを目指す。そのために、ある検索結果スポットに対して、入力した既訪問スポットがどのような関係にあるのか地図上で、一目で把握できることが望ましい。ユーザは地図上で観光スポットを検索しているものとし、まず、地図上の実際にスポットが存在する座標に検索結果スポットを配置する。次に、既訪問スポットの位置を決定する。これは検索結果スポットとの対応関係によって変化し、検索結果スポットとの類似度が高いと近く、類似度が低いと遠くなるように配置することで、意味的な位置関係を表現する。

観光スポットの対応関係の可視化を評価するために、分散表現ベクトルによる対応関係を用いた実験と相対的特徴ベクトルによる対応関係を用いた実験を行った。2つのベクトルによるそれぞれの評価実験に対して、検索対象のスポットと既訪問スポットの関連度を全体的な位置関係で表現している *Position\_Map*、位置関係による提示に加えて線の色を使って表現している *Line\_Map*、位置関係や線は用いずに *Line\_Map* で表現していたすべての情報を表形式で表現している *Table\_Map* の3つの提示手法を比較することで評価を行った。結果、分散表現ベクトルによる対応関係の可視化に関して *Position\_Map* がユーザにとってもっとも理解しやすいことがわかった。一方、相対的特徴ベクトルによる対応関係の可視化に関して *Table\_Map* がユーザにとってもっとも理解しやすいことがわかった。このことより、分散表現ベクトルを使って観光スポットの特徴抽出と対応付けを行うことによって、観光スポット間の違いが直感的に地図上に表示することができたため、一目で観光スポットの対応関係を把握できることを示した。

# 目次

第 1 章	序論	1
第 2 章	本研究のアプローチ	4
2.1	説明性向上手法	4
2.2	関連研究	6
第 3 章	ユーザの既訪問スポットの位置付けに基づく検索結果スポットの説明手法	9
3.1	概要	9
3.2	検索結果スポットと既訪問スポットの対応付け	10
3.3	対応関係を表す特徴語抽出の評価	13
3.4	観光スポットの対応付けに関する評価	15
3.5	検索結果スポット説明の有効性評価	17
3.6	まとめ	18
第 4 章	地図上における検索結果スポットの説明性向上のための観光スポットの対応関係可視化手法	20
4.1	概要	20
4.2	観光スポット間の対応関係を用いた可視化	20
4.3	対応関係の可視化評価	21
4.4	まとめ	28
第 5 章	結論	30
謝辞		32
本研究に関する発表		33
その他の発表実績		34



# 図目次

1.1	レビューサイト上で観光スポットに付与される情報例 . . . . .	2
2.1	一般的な観光スポット検索システムと提案する説明性を向上する手法 . . . . .	5
2.2	観光地検索システムの説明性向上手法 . . . . .	5
3.1	プロトタイプシステム . . . . .	10
4.1	既訪問スポットの経度座標計算の例 . . . . .	22
4.2	位置関係 ( <i>Position_Map</i> ) . . . . .	23
4.3	位置関係 + 線 ( <i>Line_Map</i> ) . . . . .	23
4.4	表形式 ( <i>Table_Map</i> ) . . . . .	23

# 表目次

3.1	既訪問スポット集合と検索結果スポット集合 . . . . .	13
3.2	対応付け結果の例 . . . . .	13
3.3	対応関係を表す特徴語の実験結果 . . . . .	14
3.4	対応付けに関する実験結果 . . . . .	16
3.5	既訪問スポットのカテゴリの同異による評価割合 . . . . .	16
3.6	システム a の方が良いと回答した被験者の自由記述 . . . . .	18
3.7	システム b の方が良いと回答した被験者の自由記述 . . . . .	18
4.1	行き先の決定しやすさの回答数割合（分散表現ベクトル） . . . . .	24
4.2	<i>Position_Map</i> と <i>Line_Map</i> の訪問履歴との関係の回答数割合（分散表現ベクトル） . . . . .	25
4.3	<i>Line_Map</i> と <i>Table_Map</i> の訪問履歴との関係の回答数割合（分散表現ベクトル） . . . . .	25
4.4	訪問先決定までにかかった体感時間の回答数割合（分散表現ベクトル） . . . . .	25
4.5	訪問先決定までにかかった労力の回答数割合（分散表現ベクトル） . . . . .	25
4.6	行き先の決定しやすさの回答数割合（相対的特徴ベクトル） . . . . .	26
4.7	<i>Position_Map</i> と <i>Line_Map</i> の訪問履歴との関係の回答数割合（相対的特徴ベクトル） . . . . .	26
4.8	<i>Line_Map</i> と <i>Table_Map</i> の訪問履歴との関係の回答数割合（相対的特徴ベクトル） . . . . .	27
4.9	訪問先決定までにかかった体感時間の回答数割合（相対的特徴ベクトル） . . . . .	27
4.10	訪問先決定までにかかった労力の回答数割合（相対的特徴ベクトル） . . . . .	27

# 第 1 章

## 序論

近年，ユーザは観光旅行をする前に Web 上の情報やガイドブックを活用して計画を立てることが多くなっている。旅行計画を立てるためには，まず訪問したいエリアを決定する。その後，観光スポット検索サイトやガイドブックの観光スポット情報を調べて観光スポットを決定する。その際，目的とする観光スポットに関する概要やアクセスするための方法などの情報を集める。

Web 上の情報に関して，観光スポットに関する総合的な情報が掲載された公式サイト，過去にそこを訪れた観光者による旅行記ブログサイト，そして，評価が記載されたレビューサイトなどのさまざまなサイトがある。しかしながら，種々のサイトが混在しており，膨大な情報から自分の嗜好に合う観光スポットの情報を得ることは簡単ではない。

ユーザの意識決定の材料の 1 つとして，るるぶトラベル<sup>\*1</sup>，Tripadvisor<sup>\*2</sup>やじゃらん<sup>\*3</sup>などの観光スポット検索サイトがある。これらのサイトには特定の観光スポットを訪問したことのあるユーザがレビューを投稿し，観光スポットに関する豊富な情報が存在している。観光スポット検索サイト上で観光スポットに付与される情報の例<sup>\*4</sup>を図 1.1 に示す。なお，これらの情報のうちユーザレビュー以外の情報（概要文・タグ・評価点など）はメタデータと呼ばれる。

観光スポット検索サイトに掲載された情報を基に，利用者（ユーザ）は自身の嗜好や目的に合致した特徴を持つ対象観光スポットを検索する。多くの観光スポット検索サイトでは，対象観光スポットに付与されるメタデータを利用した以下のような検索方法が提供されている。

- ユーザが指定したタグを持つ対象観光スポット
- 評価点などの数値メタデータに基づく並び替え

---

<sup>\*1</sup> <https://www.rurubu.travel/>

<sup>\*2</sup> <https://www.tripadvisor.com/>

<sup>\*3</sup> <https://www.jalan.net/kankou/>

<sup>\*4</sup> <https://www.jalan.net/kankou/130000/?screenId=OUW2201>



図 1.1 レビューサイト上で観光スポットに付与される情報例

- 入力単語（クエリ）を概要文中に含むオブジェクトの検索

しかし、これらのメタデータを用いた検索方法だけでは検索が困難な場合も存在する。たとえば、対象観光スポットを実際に利用することでユーザが得られる印象・感情・体験などを表す特徴に基づき対象観光スポットを検索したい場合などが該当する。また、検索できたとしても訪問したいエリア内には数多くの観光スポットが検索結果として表示される。ユーザは検索エリアに関する事前知識がないため、検索された多数の観光スポットからどの観光スポットの説明やレビューを読むべきか効率的に判断することは困難で、自身の要求に合う観光スポットを見つけることは容易ではない。

ガイドブックは、Web 上の情報に比べて整理されており、有名な観光スポットや流行のものに焦点を当て、そのスポットの料金や見どころ、そして食べ物などが掲載されている。また、ガイドブックも近年は電子化されており、技術的には検索可能である。ある観光スポットについて、複数のガイドブックが取り上げることもあり、行き方の比較検討をするためにそれらを読むのは労力を必要とする。そのため、Web 上の情報と同じ問題が起こり得る。

そこで、本研究では、観光地検索システムの検索結果である観光スポットを効率的に理解するために、ユーザが訪問したことがある観光スポットを使って説明性を向上する手法を開発する。本手法を開発するにあたり課題は 2 つ存在する。

### 1. 観光スポット間の対応付け

まず、第 1 の課題として、ユーザが訪問したことがある観光スポットと、検索結果スポットをどのように対応付けると説明として効果的なのか明らかではない。ある観光

スポットに対応する観光スポットを判定する手法として、レビュー やカテゴリが類似したコンテンツであるという観点で判定する方法が一般的である。カテゴリなどのメタデータが類似する対応付けの場合には、大まかな利用目的レベルでの説明が可能であると考えられる。レビューが類似するという場合には、より細かい体験のレベルでの説明が可能となると考える。このように対応付けの方法によって、その説明性も変化すると考えられる。本研究では、前述の2つに、さらに、相対的な特徴による類似という、より詳細なレベルでの対応付けを加え、比較することで、説明に適した対応付け方法を明らかにする。

## 2. 地図上における観光スポットの対応関係の可視化

第2の課題は、数多くあるユーザの過去の訪問先と検索結果をどのように提示すると、説明として適切なのか明らかではないことである。まず、地理的情報の可視化と意味的な関係を同時に可視化することを考える。地理情報の可視化としては、抽出した情報を地図上にマッピングするのが一般的である。また、意味的な関係性の可視化として、グラフモデル上でオブジェクト間のつながりを関連度で表現することが多い。観光スポット間の意味的な対応関係を可視化するために、検索結果スポットは地図上で実在する座標に固定し、既訪問スポットを関連度に基づいて適切に配置することが妥当であると考える。そこで、ある検索結果スポットに対して、入力した既訪問スポットがどのような関係にあるのか地図上で、一目で把握できるように、既訪問スポットと検索結果スポットの対応関係を可視化する手法を提案し、一般的な表形式や表示情報を増加させた形式と比較する。このことにより、説明として適切な可視化方法を明らかにする。

以下に本論文の構成を示す。まず、2章で本研究のアプローチについて述べる。ここでは観光支援システムのアルゴリズムと関連研究について述べる。この観光支援システムのための2つの観点に基づく各手法は3章と4章で述べる。3章では、ユーザの既訪問スポットの位置付けに基づく検索結果スポットの説明手法について述べる。4章では、地図上における検索結果スポットの説明向上のための観光スポットの対応付け可視化手法について述べる。5章では、まとめと今後の課題を述べる。

## 第2章

# 本研究のアプローチ

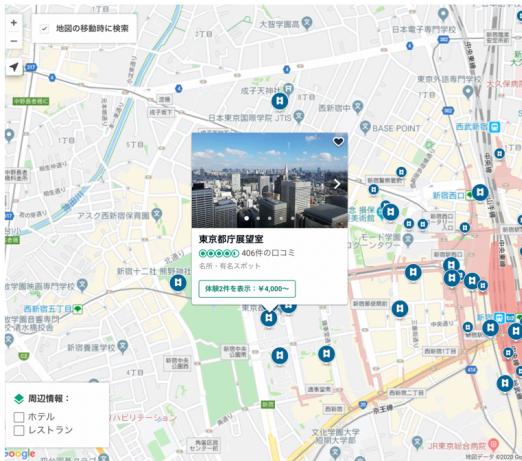
### 2.1 説明性向上手法

一般的な観光スポット検索システムは、訪問したいカテゴリやエリアを入力して検索を行う。検索結果は観光スポットのリストの形式としてユーザに表示される。また、地図上にプロットして表示する機能を備えているものもある。地図上の検索としては、Google Mapなどの地図サイドで「観光」をキーワードに検索することも可能である。このように、検索結果の一貫性や網羅性は高い。図2.1の左側は TripAdvisorによる西新宿エリアの地図表示であり、観光スポットが地図上にプロットされた検索結果が表示される。一方、検索結果の詳細は、そのスポットをクリックして詳細情報を表示するまでわからない。このとき地図上に、自分が訪れたことがあるスポットと検索結果の関係がわかる情報が表示されていれば、詳細情報を確認する前に、そのスポットについて知ることができると考えた。このような、検索結果自体が理解しやすくなることを説明性の向上と定義する。図2.1の右側は左側の検索結果に説明性向上を施したものである。図2.1の左側は「東京都庁展望室」などの検索結果スポットのみ地図上で表示されているに対し、右側は検索結果スポットと既訪問スポットが表示されている。図2.1の右側の図から一目で、「梅田スカイビル」は「東京都庁展望室」との関連性が強い、「住吉大社」は関連性が低いことがわかる。

観光支援システムの1つに、求める観光スポット情報を発見するために観光スポットの理解をしやすくすることが考えられる。観光情報収集までの過程を効率化することで、ユーザの観光計画の敷居は下がり、訪れたことない観光スポットの理解の向上も望める。本研究では、ユーザ今まで体験したことがありかつ気に入った観光スポットを利用し、訪問したいエリア内の観光スポットに対して関連付けすることで、手軽に観光情報を理解する説明性向上手法を提案する。観光地検索システムの説明性向上手法の流れを図2.2に沿って説明する。

まずははじめに、ユーザは体験履歴の中で今まで訪れたことがあり、かつ気に入った観光スポットを入力する。次に、これから行ってみたい都道府県やエリアを入力する。これらの

## 一般的な観光スポット検索システム



## 提案する説明性を向上する手法



図 2.1 一般的な観光スポット検索システムと提案する説明性を向上する手法

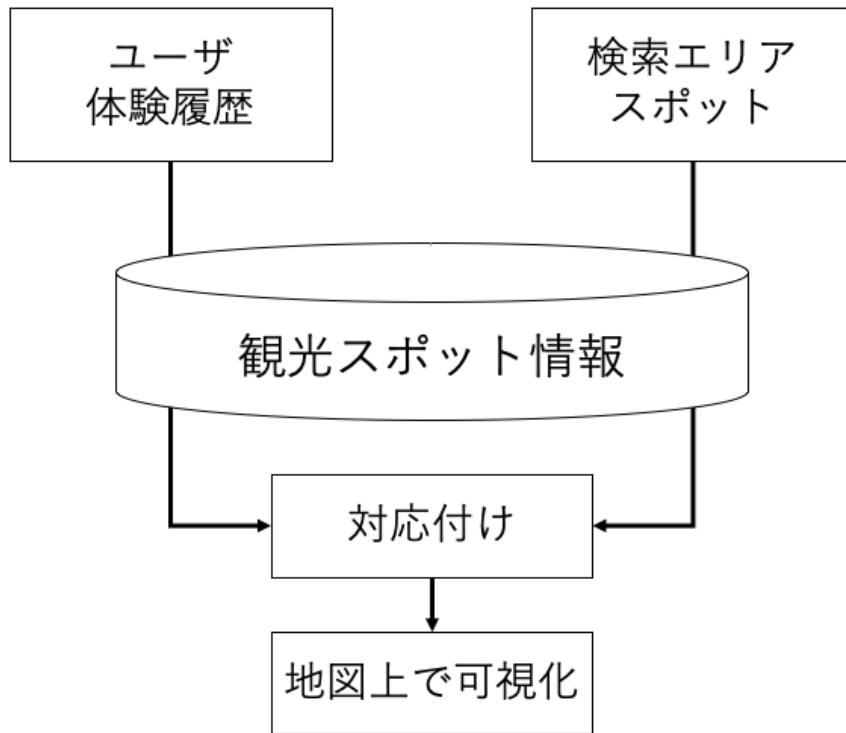


図 2.2 観光地検索システムの説明性向上手法

入力に対し、ユーザの体験履歴と検索結果を対応付ける。すなわち、ユーザの未知な観光スポットに対する理解を支援するため、すでに訪問したことがある観光スポットの特徴を用いて、未訪問エリアの観光スポットを説明するための対応付けを行う。

本研究における観光スポットの対応付けに関して、ユーザによって、その強調するべき特徴が変化することがあり得る。たとえば、日本の有名な観光スポットを履歴に持つユーザに

とっての「東京都庁舎展望台」と「金閣寺」について考える。このとき、このユーザにとっての「金閣寺」の特徴は、お寺、金色、京都などである。また、「東京都庁舎展望台」の特徴は、展望、夜景、建物、新宿などが考えられる。一方、京都の寺院を履歴に持つユーザにとっての「金閣寺」と「清水寺」について考える。このとき、同じ「金閣寺」でもお寺や京都などは特徴にはならず、このユーザにとっての「金閣寺」の特徴は、金色、金箔、輝きなどとなる。また、「清水寺」の特徴は、舞台や一望などである。

対応付けを行ったあと、ユーザに検索結果スポットを説明するための観光スポット特徴語を抽出する。最後に、地図上で既訪問スポットと検索結果スポットの対応関係と特徴語を可視化することで、ユーザが検索結果スポットに対する理解を容易化することを目指す。

## 2.2 関連研究

本研究では、ユーザの体験履歴の中のレビューを分析し、地図上で対応関係を可視化して観光情報を理解する説明性向上手法を提案する。本研究の特徴は、体験履歴を検索結果の説明性向上に用いることと、地図上で意味的な関係性を可視化することである。そこで、従来のユーザの体験履歴を用いた検索や推薦に関する研究を紹介することでアプローチの違いを述べる。次に、地理的な情報の可視化と意味的な関係の可視化に関する研究を紹介することで可視化の問題を整理する。

### 2.2.1 体験履歴による検索や推薦システムに関する研究

本研究では、ユーザの体験履歴を検索結果の個人化した説明に用いる。そこで、検索や推薦システムでの体験履歴の一般的な用いられ方を紹介することでアプローチの違いを説明する。

まず、体験履歴は個人化のために用いられることが多い。倉島ら[?]は、Flickrに投稿された写真のジオタグ情報を人々の旅行履歴として利用した旅行ルート推薦手法を提案した。この手法では、ユーザの現在地から行きやすい場所とユーザの興味に合致した場所に移動しやすいと仮定し、行動モデルを生成している。ユーザのジオタグ付き写真集合は、時間情報でソートすると個人の旅行履歴とみなすことができると考え、ジオタグ情報を利用してユーザの行動モデルを生成している。Chengら[?]は、自由に利用できるコミュニティ投稿の写真を活用して、パーソナライズされた旅行のおすすめに焦点を当て、特定のユーザプロファイルまたは属性を考慮し、パーソナライズされた旅行の推薦を行うことを提案した。このように、個人化のために履歴から何らかのユーザモデルを抽出し、そのユーザモデルで検索や推薦を行う。一方、本研究では、体験履歴そのものをそのまま説明として用いる。発展とし

て、履歴から抽象化した嗜好を抽出し、説明として用いることも考えられるが、ユーザが理解しやすいことを重視するため、複雑なユーザモデルを必要としない。

体験履歴の用い方として、検索対象の特徴量とするアプローチがある。観光地検索するとき、松本ら [?] はクチコミから特徴語を抽出して利用する研究を行った。抽出対象を任意の名詞として、4種類の手法、TF-IDF、ATF(Average Term Frequency)、ポアソン確率、エントロピーのうちどの手法が特徴語抽出に適しているのか検討を行った。また、抽出した特徴語を利用した検索支援システムを試作し、実験を通して特徴語提示の効果を検証した。嶋田ら [?] は Web から観光情報を抽出し、複数の特徴ベクトルから観光地間の類似性を評価することで、観光地を推薦するシステムを提案した。観光地の特徴ベクトルは、知恵袋・ブログ上での共起キーワードと時系列分布、知恵袋上でのカテゴリ構造、観光地周辺施設、地図画像から生成した。これらの特徴ベクトルから観光地間の類似度の測定を行い、類似度の高い観光地を推薦した。廣島ら [?] は地理情報検索の際のクエリ入力支援として、提示する特徴語の抽出手法について研究を行った。この手法では、各ブログ記事から特徴語候補の抽出および地点の特定を行った。具体的には、特徴語の候補を Wikipedia の見出し語に限定し、ポアソン確率を用いて特徴語抽出を行った。また、地点の特定のための地名表現の抽出および緯度経度への変換の方法としては平野ら [?] の手法を用いた。野守ら [?] は日本全国の観光地のクチコミデータを用いて、観光客が話題にする観光テーマを確率的に抽出し、そのテーマを軸として各観光地の特徴を定量的に評価した。また、クチコミのテキストデータにテキストマイニングを実行して表現を抽出し、観光地ごとにその表現の出現頻度を集計したクロス集計表に PLSA を実行することで、観光客のクチコミだけに基づいた観光テーマの抽出と観光地の特徴分析を行なった。本研究においても観光スポットの特徴ベクトルの生成に、全ユーザのレビューを用いており、この部分において、これらのアプローチと同様である。

### 2.2.2 情報可視化に関する研究

本研究では、地理的な情報の可視化と意味的な関係の可視化に取り込む。地理的な情報の可視化としては、抽出した情報を地図上にマッピングするのが一般的であり、本研究もそれに従っている。代表的な研究を以下に紹介する。櫻川ら [?] は、ソーシャルメディア上にアップロードされた写真のジオタグ情報と撮影時刻に基づいて写真の撮影者を分類する手法を提案した。分類された撮影者（観光者、在住者）ごとにホットスポットを可視化した。また、ジオタグ情報と撮影時刻以外にテキストタグを加えて、写真が撮影された地域で行われるイベントの穴場スポットを発見し、地図上に提示する手法を提案した [?]. 倉田ら [?] は、新しい観光情報サービスの形として「他の旅行者がどこで撮影したいか」というデータから

観光ポテンシャルを可視化し、地図上に表示する方法を提案した。

意味的な関係性の可視化としては、グラフモデル上でオブジェクト間のつながりを関連度を表現することが多い。代表的な研究を以下に紹介する。Kitamura ら [?] は、一般的な物体認識を用いて、過去の個人旅行写真から推定したユーザの旅行の嗜好に基づき観光地を推薦する方法を提案した。物体認識システムを用いて、写真で撮った被写体情報のキーワードを取得し、グラフ視覚化技術によってキーワードの共起を表現した。また、グラフの視覚化技術に基づいて旅行写真付きのグラフを視覚化するユーザインターフェイスを紹介した。上村ら [?] は、ユーザが投稿したタグ付き画像を用いたファッショスタイルの関係性の可視化手法を提案した。ファッショスタイルは類似するタグを空間座標に固定することによって関係を表している。本研究では、検索結果スポットは地図上で実在する座標に固定され、配置の自由があるのは既訪問スポットのみという制約があり、この問題に取り組む。

## 第3章

# ユーザの既訪問スポットの位置付けに基づく検索結果スポットの説明手法

本章では、ユーザの未知なスポットに対する理解を支援するため、すでに訪問したことがある観光スポットの特徴を用いて、未訪問エリアの観光スポットを説明する手法を述べる。また、対応関係を表す特徴語抽出、観光スポットの対応付け、検索結果スポットの説明の有効性についての評価実験を行う。

### 3.1 概要

ユーザは検索エリアに関する事前知識がないため、どのスポットのレビューを読むべきか効率的に判断することは困難である。さまざまな観光スポットを効果的に理解するためには、ユーザが訪問した経験のあるスポットを使って検索結果スポットと類推できることが重要であると考えた。たとえば、日本に初めて訪れるフランス人旅行者に対し、検索結果スポットである東京の「表参道」をパリにおける「シャンゼリゼ通り」と表現すると理解しやすいであろう。この考え方は、ユーザが以前に経験した物事を現在の物事に適用する一種の類推である。既知の知識（ベースと呼ぶ）から概念（ターゲットと呼ぶ）を獲得するときに類推思考が働くとされる[?]. 構造の類似性には3種類あり、特徴の共有数で決まる「対象レベルの類似性」、ベースに存在する関係とターゲットに存在する関係の共有度に基づく「関係レベルの類似性」、および題の解法あるいは目標レベルでの類似性である「プラグマティックな類似性」とがある[?], [?]. 本研究では、類推の質を明示的に扱うため、「関係レベルの類似性」に近いと考えられる。

図3.1は、構築したプロトタイプシステムであり、訪問履歴として金閣寺等の京都のスポットを持つユーザが東京のエリアを検索した画面である。未訪問エリアにある迎賓館に対して金閣寺が豪華絢爛という観点で類似する様子を示している。このように、抽出したキーワードを提示することで、ユーザの検索結果スポットに対する理解の支援を目指す。



図 3.1 プロトタイプシステム

### 3.2 検索結果スポットと既訪問スポットの対応付け

検索結果スポットを理解しやすくするために、最も位置付けの近いユーザの既訪問スポットを対応付ける手法を提案する。まず、ユーザがすでに訪問した複数個の観光スポットとこれから訪問したい観光スポットエリア情報を入力する。本手法では、観光スポットのユーザレビューを用いて特徴ベクトルを生成する。この特徴ベクトルには、カテゴリなどのメタデータによるもの、そのスポットのレビュー文から生成した文書ベクトルが一般的である。本研究では、さらに、ユーザの観光地の位置付けを抽出するために、あるスポットに対し他の訪れたスポットと比較した相対的特徴ベクトルを用いる。同様に、検索結果スポットもエリア内の各スポットの特徴ベクトルを求める。また、未訪問エリアの各観光スポットの特徴ベクトルを、そのエリアの他の観光スポットと比較して計算する。次に、特徴ベクトルの類似度によって既訪問スポットを検索結果スポットと対応付ける。最後に、その関係性を説明するためのキーワードを抽出する。

なお、篠田ら[?], 藤田ら[?]の研究にあるようにGPSなどの位置探知デバイスから自動的にユーザの移動情報を利用して、移動の特徴を分析し興味のあるスポットを取り出す技術もある。本研究では、ユーザの体験履歴の中で今まで訪れたことがある、かつ気に入った観光スポットを入力することとしているが、このような、位置探知デバイスの技術を使って体験履歴を取得することが可能と考えられる。

### 3.2.1 観光スポットのユーザレビューを用いた特徴ベクトル生成と対応付け

本研究では、スポットの特徴ベクトル生成に Paragraph Vector モデルを用いる。Paragraph Vector モデルは、Mikolov ら [?] によって考案された単語の特徴ベクトル学習手法である word2vec を拡張し、Le ら [?] によって提案された文書の特徴ベクトル学習手法である。Paragraph Vector モデルでは Harris の分布仮説 [?]<sup>\*1</sup>に基づき、「ある文書中である単語列を与えられたとき、次に出現する単語を予測する」というタスクをニューラルネットワークに学習させることで、文脈や単語の語順を考慮した文書の特徴ベクトルを生成することができる。Paragraph Vector モデルでは、生成した文書の特徴ベクトルを文書の分散表現と呼ぶ。

分散表現を用いて観光スポットの特徴ベクトルを作成する。このとき、観光スポット毎のレビューをまとめて1つの文書として扱う。本研究では、分散表現を計算するために Python のライブラリである gensim<sup>\*2</sup>を利用する。学習方法として、Distributed Bag-of-Words を利用して、各スポットの全レビューを使って300次元で作成したベクトルを使う。既訪問スポットや検索結果スポットのレビューべクトルは、形態素解析器である MeCab[?] に辞書「mecab-ipadic-NEologd」<sup>\*3</sup>用いて、分かち書きし原形にしたレビューを利用して作成する。このようにしてできた一般的な文書ベクトルを本論文では、分散表現ベクトルと呼ぶ。

それに対し、本研究における観光スポットの対応付けに関して、ユーザによって、その強調るべき特徴が変化することがあり得る。たとえば、日本の有名な観光スポットを履歴に持つユーザにとっての「東京都庁舎展望台」と「金閣寺」について考える。このとき、このユーザにとっての「金閣寺」の特徴は、お寺、金色、京都などである。また、「東京都庁舎展望台」の特徴は、展望、夜景、建物、新宿などが考えられる。一方、京都の寺院を履歴に持つユーザにとっての「金閣寺」と「清水寺」について考える。このとき、同じ「金閣寺」でもお寺や京都などは特徴にはならず、このユーザにとっての「金閣寺」の特徴は、金色、金箔、輝きなどとなる。また、「清水寺」の特徴は、舞台や一望などである。このような特徴を相対的特徴と定義する。あるスポットに対し他の訪れたスポットと比較した相対的特徴ベクトル  $r_{state,i}$  を、式3.1として定義する。

$$r_{state,i} = s_i - \text{average}(\{s | s \in S_{state}, s \neq s_i\}) \quad (3.1)$$

相対的特徴ベクトルは、そのスポット自身の分散表現ベクトルから他のスポットの分散表現ベクトルの平均を引いた値によって得られる。 $S_{state} = \{s_1, s_2, \dots, s_n\}$  は、既訪問スポットもしくは検索結果スポットのベクトル集合である。 $state$  が  $f$  であれば、既訪問スポット集

---

<sup>\*1</sup> Harris の分布仮説は単語の分散モデルでも使われており、「同じ文脈で出現する単語は類似した意味を持つ」という仮説である。

<sup>\*2</sup> <https://radimrehurek.com/gensim/models/doc2vec.html>

<sup>\*3</sup> <https://github.com/neologd/mecab-ipadic-neologd/>

合を意味し,  $state$  が  $u$  であれば, 検索結果スポット集合を意味する.  $s_i$  は集合  $S_{state}$  内の観光スポット  $i$  の分散表現ベクトルを示している.

未訪問エリア内のスポットを既訪問スポットと対応付ける. 具体的には, ある検索結果スポットの特徴ベクトルに最も類似度が高い既訪問スポットを対応付ける. 類似度計算には, コサイン尺度を用いる. ただし, 類似度が閾値以下である場合は対応付けを行わない.

カテゴリなどのメタデータによる対応付け手法を説明する. この対応付けはここまでと異なり, 特徴ベクトルではなく, 観光スポット検索サイトでスポットを検索するためのメタデータであるカテゴリ, 滞在時間, 訪問時期の一致を用いる. まず, 検索結果スポットに対し, すべてが一致する既訪問スポットを抽出する. 抽出できない場合は, 季節, 滞在時間, カテゴリの順に条件を緩和する. 既訪問スポットが複数ある場合は, レビュー数が最も多いスポットを選択する. メタデータの例を以下に示す.

- カテゴリ : 神社・神宮・寺院, 観光施設・名所巡り等
- 滞在時間 : 1 時間未満, 1–2 時間等
- 訪問時期 : 1–12 月, 春, 夏, 秋, 冬

### 3.2.2 対応する観光スポットの特徴語抽出

観光スポットの対応関係を示すだけでは, どのような点で対応するのかを理解するのは難しい. そこで, 検索結果スポットと既訪問スポットの関係性を表すキーワードをユーザに提示する. しかし, メタデータそのものや, 分散表現である特徴ベクトルから単語の特徴を得ることはできないため, 他の方法を用いて単語を抽出する.

すべてのレビューを 3.2.1 節と同様に形態素解析器 MeCab によって単語に分割する. ただし, 助詞, 助動詞, 連体詞, 記号, ストップワードを削除している.

キーワード抽出手順について説明する. まず, TF-IDF 法を使って対象となる既訪問スポットと検索結果スポットの特徴語と TF-IDF 値を求める. IDF 値 [?] を算出する集合は  $S_{state}$  を用いる.

2 つのスポットに共通する特徴語の TF-IDF 値の調和平均を用いて, 対応付けしたスポットの説明可能なキーワードとして抽出する. 各単語のスコアを式 3.2 によって定義する.

$$score(t, s_f, s_u) = \frac{2 \times tfidf(t, s_f) \times tfidf(t, s_u)}{tfidf(t, s_f) + tfidf(t, s_u)} \quad (3.2)$$

$tfidf(t, s_f)$  と  $tfidf(t, s_u)$  は同じ単語に関する既訪問スポット  $s_f$  における TF-IDF 値と検索結果スポット  $s_u$  における TF-IDF 値を示している. 単語スコアの上位  $N$  個の単語を説明情報としてユーザに提示する.

表 3.1 既訪問スポット集合と検索結果スポット集合

既訪問スポット名	検索結果スポット名
浅草寺	東京ディズニーランド (R)
小田原城址公園	新宿御苑
伏見稻荷大社	東京スカイツリー
奈良公園	東京タワー大展望台
三島スカイウォーク	明治神宮

表 3.2 対応付け結果の例

検索結果スポット	既訪問スポット	特徴語
新宿御苑	小田原城址公園	お花見, 咲き誇る, 園内, 桜の時, のんびり, 手入れ, 自然, 遊具, ツツジ
東京スカイツリー	三島スカイウォーク	富士山, 摆れ, 高所恐怖症, 摆れる, 天井, 絶景, エレベーター, パノラマ, 展望デッキ, 昇る

### 3.2.3 実行例

表 3.1 に示す、既訪問スポット集合と検索結果スポット集合を用いて相対的特徴ベクトルを用いて対応付けを行った結果を説明する。検索結果スポットは首都圏内を想定し 5 つのスポットを選んだ。表 3.2 は、3.2 節で述べた提案手法を用いて抽出した対応付いたスポットの特徴語を示している。

検索結果スポット「新宿御苑」に対し、既訪問スポット「小田原城址公園」が対応付いた。他に候補として「奈良公園」が考えられるが、「小田原城址公園」には花や遊具に対する記述が多く、「奈良公園」には鹿や草に対する記述が多い。「新宿御苑」には花や遊具に対する記述が多いため、「小田原城址公園」と適切に対応付けたと考える。

検索結果スポット「東京スカイツリー」に対し「三島スカイウォーク」が対応づけられた。どちらも眺めや高さに特徴があり関連付けられたと考えられる。また、特徴語を見ると「富士山」が抽出されており、見える景色を想起させるような対応付けを行っていると考える。

## 3.3 対応関係を表す特徴語抽出の評価

### 3.3.1 実験内容

本実験では、2016 年 9 月末までのじゃらん<sup>\*4</sup>から得られたレビューデータを用いる。また、本実験において類似度を計算するときの閾値を 0.125 とする。これは、以降の 3.4 節、

<sup>\*4</sup> <https://www.jalan.net/kankou/>

表 3.3 対応関係を表す特徴語の実験結果

選択肢	スポット関係	キーワード	算術平均法	乗算法	調和平均法
1	あり	明確になった	28.28%	31.31%	29.29%
2	あり	初めて気づいた	35.35%	31.31%	35.35%
3	あり	ない	10.10%	14.14%	12.12%
4	ない	—	26.26%	23.23%	23.23%

3.5 節の実験でも同様である。提案手法のうちの特徴語抽出部分について評価を行うため、他のキーワード抽出手法と比較して評価を行う。なお、特徴ベクトルは相対的特徴ベクトルに固定している。以下の 3 つの方法を比較する。

- 算術平均法
- 乗算法
- 調和平均法

キーワードを抽出するとき、算術平均法では、2 つのスポットに共通する特徴語のスコアとして算術平均を計算する。抽出した単語のスコアを式 3.3 によって定義する。

$$score(t, s_f, s_u) = \frac{tfidf(t, s_f) + tfidf(t, s_u)}{2} \quad (3.3)$$

乗算法では、2 つのスポットに共通する特徴語のスコアとして乗算を計算する。抽出した単語のスコアを式 3.4 によって定義する。

$$score(t, s_f, s_u) = tfidf(t, s_f) \times tfidf(t, s_u) \quad (3.4)$$

### 3.3.2 実験手順

クラウドソーシングのサービスである、 CrowdWorks<sup>\*5</sup>を利用して 23 人の被験者を集めた。まず、被験者は 4 個から 10 個の自分がすでに訪れたことのある観光スポットを入力した。次に、被験者は旅行などで行ったことがなく、これから訪れたい都道府県やエリアを入力した。

本実験では、それぞれの手法において、対応付けられた既訪問スポットと検索結果スポット、およびスポット同士の関係を説明する特徴語を最大で 5 つまで提示した。結果を評価するために、被験者は以下の 4 つの選択肢から 1 つを選んだ。

1. 2 つのスポットには関係性があり、キーワード<sup>\*6</sup>によって関係が明確になった。

---

<sup>\*5</sup> <https://crowdworks.jp/>

<sup>\*6</sup> 関係を説明する特徴語のことであるが、被験者には単にキーワードと示した。

2. 2つのスポットの関係性があり、キーワードによって初めて気がついた。
3. 2つのスポットには関係性があるが、キーワードは関係を表していない。
4. 2つのスポットには関係性がない。

### 3.3.3 実験結果と考察

表3.3は、それぞれの手法における選択肢1から4それぞれの選択割合を示している。乗算法では、多くの被験者が選択肢1を選択している。このことから、乗算法は明らかな関係を表現する特徴語抽出ができることがわかった。算術平均法と調和平均法では、選択肢1の割合が減少し、選択肢2の割合が増加している。このことから、算術平均法と調和平均法は隠れた関係を表現する特徴語を見つけて被験者に提示することができると考えた。

しかし、算術平均法では、選択肢4の割合が増加しているため、調和平均法より無関係な特徴語を抽出しやすくなったといえる。また、選択肢1と2を合計すると、調和平均法が最も高くなっている。このことから、スポットの関連を説明する特徴語抽出手法としては調和平均法が妥当といえる。

## 3.4 観光スポットの対応付けに関する評価

### 3.4.1 実験内容

特徴語抽出については3.2.2節の手法を用いる。提案手法のうちの対応付け部分について、3.2.1節で説明した以下の3つの手法を比較して評価を行う。

- メタデータ法
- 分散表現法
- 相対的特徴法

クラウドソーシングのサービスである、CrowdWorksを利用して24人の被験者を集めた。実験手順は3.3.2節で説明した手順と同様である。

### 3.4.2 実験結果と考察

対応付けしたスポットの合計数は285である。表3.4は、それぞれの方法における選択肢1から5それぞれの選択割合を示している。メタデータ法について、多くの被験者が選択肢2を選択している。このことから、メタデータ法は明らかな関係にあるスポット同士を関連付けることができる事がわかった。他の方法と比較して、相対的特徴法では、選択肢1の割合が減少し、選択肢2の割合が増加している。このことから、相対的特徴法は隠れた関係

表 3.4 対応付けに関する実験結果

選択肢	スポット関係	キーワード	メタデータ法	分散表現法	相対的特徴法
1	あり	明確になった	41.30%	33.85%	29.36%
2	あり	初めて気づいた	43.48%	47.69%	48.62%
3	あり	ない	2.17%	2.31%	2.75%
4	ない	一	13.04%	16.15%	19.27%
対応付け数			46	130	109

表 3.5 既訪問スポットのカテゴリの同異による評価割合

	選択肢	分散表現法	相対的特徴法
異なる場合	1	19.23%	21.10%
	2	24.62%	25.69%
同じ場合	1	14.62%	8.26%
	2	23.08%	22.94%

を見つけて被験者に提示することができると考えた。しかし、選択肢 4 の割合も増加しているため、無関係なスポットを抽出しやすくなつたといえる。

選択肢 1 と 2 を合計すると、メタデータ法が最も高くなる。このことから、説明するためのスポットの対応付けの精度が最も高いといえる。しかし、メタデータ法は対応付けすることができるスポットの数が他の手法に比べ半分以下であり、多くの検索結果スポットに説明を付与することができない。

分散表現法と相対的特徴法を詳細に考察するために、既訪問スポットの入力別に分析した。入力した既訪問スポットの半数以上が同一カテゴリのスポットであった被験者（表中の「同じ場合」）とそうではない被験者（表中の「異なる場合」）に分けて集計した。表 3.5 は選択肢 1 と選択肢 2 の評価の割合である。相対的特徴法を使用すると、カテゴリが異なる場合において、分散表現法よりも選択肢 1 および選択肢 2 の割合を増加させることができる。相対的特徴ベクトルを用いることによって、カテゴリをこえて各スポットの特徴を求めることができるといえる。カテゴリが同じ場合の選択肢 1 は相対的特徴法では、少なくなっている。これは、相対的特徴ベクトルとしては、カテゴリの特徴が打ち消されたベクトルが生成されるため、明らかな関係となる対応付けが減ったと考えられる。しかし、選択肢 2 が増えずに選択肢 4 が増え、無関係に見える対応付けを増やす結果となつた。そのため、入力スポットの種類によって対応付け手法を切り変えることが有効と考えられる。

### 3.5 検索結果スポット説明の有効性評価

#### 3.5.1 実験内容

提案手法の有効性を評価するために以下の 2 つのシステムを使って比較する。提案手法では、対応付けに相対的特徴法、キーワード抽出に調和平均法を用いた。

- a. 検索結果スポット名のみを表示
- b. 検索結果スポット名、対応する入力スポット、対応を説明する特徴語を表示（提案手法）

被験者の入力は、これまでの実験と同様に、4 個から 10 個の自分がすでに訪れたことのある観光スポットと訪問したいエリアである。ただし、エリアは 2 つ入力し、それぞれ別々にシステム a、システム b の入力として用いた。なお、システム a とシステム b はランダムな順番で実行される。システムを評価するために、被験者は以下の 2 つの設問について回答した。また、それらの選択理由について自由記述で答えた。

Q1. 表示されたスポットの詳細情報はどちらの方が分かりやすいかを選択してください。

Q2. 旅行の計画を立てる際にどちらの方が使いたいかを選択してください。

被験者はクラウドソーシングのサービスである、CrowdWorks を利用して集めた 50 人を用いた。

#### 3.5.2 実験結果と考察

Q1 に対する回答について、システム a とシステム b の結果はそれぞれ 12 件と 38 件となった。被験者の回答理由の例を表 3.6（システム a）および、表 3.7（システム b）に示す。また、旅行の計画を立てる際にどちらの方が使いたいかに対する回答について、システム a とシステム b の結果はそれぞれ 10 件と 40 件となった。被験者の回答より、得られた 2 つのシステムの特徴をまとめた。

システム b では、キーワードや関連している情報が表示されており分かりやすいという特徴がある。これに対し、システム a では、スポット名のみが表示されているためシンプルで分かりやすいという特徴があるが、なぜそのスポットが検索されたのか分からぬなどの否定的な回答も多数あった。システムの表示情報が「シンプルな方が良い」と「詳細な情報が良い」の 2 つに大きく分けられているが、Q2 の回答より 80% の被験者はシステム b の方が

表 3.6 システム a の方が良いと回答した被験者の自由記述

スポット名だけ書いてあるのでごちゃごちゃにならず、一目でどこに何があるか分かるからです。
スポットだけ表示されるところがシンプルで分かりやすかったから。
シンプルにそのエリアの観光スポットを知ることが出来る。

表 3.7 システム b の方が良いと回答した被験者の自由記述

自分が訪れた事のあるスポットと関連性があり連想しやすいから。
どのような場所か想像しやすいので、計画を立てるのにも使いたいと思ったから。
何と関連して表示された等詳細な情報が表示されわかりやすかった。

使いたいと回答しており、提案手法の方が評価が高いといえる。また、システム b に関してスポットの写真を載せるとより分かりやすいなどの意見があり、画像による関連性の説明という拡張が考えられる。

### 3.6 まとめ

観光スポットを効果的に理解するために、観光スポット間の対応付けを行い、検索結果スポット説明の有効性を評価するための実験を行った。検索結果スポットを簡単に理解するためには、ユーザにどのような情報を提示することによって理解につながるかが重要である。従来では、カテゴリ、滞在時間などのメタデータを使って観光スポットの検索を行っている。検索したあとに、検索結果スポットをひとつずつ見ていくことが多くなっている。これでは、ユーザにとって理解するために時間と労力が必要となり、検索結果スポットを効果的に理解することにつがらない。検索結果スポットを効果的に理解するために、ユーザに今まで訪問したことのある既訪問スポットを用いて検索結果スポットを説明するのが効果的であろうと考えられる。また、検索結果スポットを説明するためにその特徴語を表示することが効果的と考えられる。対応付けによる説明を効果的にするために、既訪問スポットと検索結果スポットの対応付けはレビューを用いて、相対的特徴ベクトルによる類似により詳細な対応付けを行う。

3.3 節では、相対的特徴ベクトルで対応付けした既訪問スポットと検索結果スポットの特徴語を抽出した実験を行って、調和平均の計算による単語抽出の方が妥当である。3.4 節では、メタデータ法、分散表現法、相対的特徴法の 3 つの手法で対応付けの実験を行った。ユーザに提示する情報は 3.3 節で確認した調和平均によって抽出した特徴語である。

メタデータでは、速く観光スポットの検索ができるが、絞り込みによって検索される結果

が少なくなる。メタデータに対して、分散表現と提案手法では対応付けするために処理する時間がかかる。また、メタデータによる検索結果スポットはユーザが設定した大きな利用目的内にある。ユーザがこれから訪問したいエリアのカテゴリに収まることが多いため、利用目的が明確な場合においてメタデータ法はユーザにとって理解しやすい。

提案手法において、相対的特徴ベクトルを利用した場合、カテゴリの特徴が打ち消されたベクトルが生成され、検索結果スポットはユーザが設定したカテゴリ情報だけでなく、より詳細な特徴を持つ検索結果スポットの特徴語が提示される。提案手法によって提示した検索結果スポットの特徴語は、ユーザにとって新しい説明情報としてスポットの理解につながる。

観光スポットを効果的に理解する点において、ユーザはこれから訪問したいエリアに対して、カテゴリがすでに決まっている場合では、メタデータ法で検索された観光スポットが効果的に理解しやすいと考えられる。それに対し、ユーザは訪問したいエリアのスポットに対してより詳細な情報を知りたい場合、相対的特徴ベクトルで対応付けしたスポット間の特徴語を提示することによって効果的に理解しやすいと考えられる。

観光スポットを効果的に理解するために、ユーザに検索結果スポット名だけではなく、対応する既訪問スポットとそれを説明する特徴語を加えることが理解につながると考えられる。地図上では一般的な検索結果はスポット名や場所情報がユーザに表示されるが、3.5節では、対応する既訪問スポットと説明するための特徴語を提示することで説明の有効性評価を行った。特徴語と対応する既訪問スポットを提示することによって、検索結果スポットとの関連性が連想しやすくなって、観光スポットを効果的に理解することにつながる。

3.3節、3.4節、3.5節で行った3つの評価実験を通して相対的特徴ベクトルを利用して、対応付けした方がより観光スポットの詳細を理解することができる。

## 第 4 章

# 地図上における検索結果スポットの説明性向上のための観光スポットの対応関係可視化手法

本章では、観光スポット間の対応関係を地図上に可視化する手法を提案する。また、プロトタイプシステムを構築し、既訪問スポットと検索結果スポットの説明性向上のための観光スポットの対応関係を地図上に可視化した効果を評価する実験を行う。

### 4.1 概要

近年、ユーザは Web 上の観光情報を活用して旅行計画を立てることが多くなっている。しかし、旅行は一般的に訪れたことがないスポットに行くことが多いため、観光情報を適切に理解することは困難である。そこで、訪問したことがある観光スポットの特徴を用いて、未訪問エリアの観光スポットを対応付けして、ユーザの未知なスポットに対する理解を支援することを考える。観光スポット間の対応関係を地図上に可視化する手法を提案する。

本手法では、まず、3 章で述べた手法により、スポットの特徴ベクトル間の類似度に基づいて既訪問スポットと検索結果スポットを対応付けし、その対応関係を説明するための特徴語を抽出する。次に、地図上で表示するための意味的な位置関係を決定する。具体的には、検索結果スポットからの地理的な距離と類似度によって存在度を決定して、意味的な位置関係に基づいて観光スポットの対応関係をわかりやすくするための既訪問スポットの座標を算出する。また、プロトタイプシステムを構築し、既訪問スポットと検索結果スポットの説明性向上のための観光スポットの対応関係を地図上に可視化した効果を評価する実験を行う。

### 4.2 観光スポット間の対応関係を用いた可視化

検索結果スポットを理解しやすくするために、もっとも位置付けの近いユーザの既訪問スポットを対応付ける手法を提案した。この手法では、3 章で述べた対応付けを前提としてい

る。簡単に処理を説明する。ユーザがすでに訪問した複数個の観光スポットとこれから訪問したい観光エリアを入力とする。ユーザが入力した観光スポット集合のユーザレビューを用いてそれぞれのスポットの特徴ベクトルを生成する。同様に、検索結果スポットもエリア内の各スポットの特徴ベクトルを求める。次に、特徴ベクトルのコサイン類似度によって既訪問スポットを検索結果スポットと対応付ける。最後に、その関係性を説明するためのキーワードをTF-IDF特微量と調和平均で抽出する。

ある検索結果スポットに対して、入力した既訪問スポットがどのような関係にあるのか地図上で、一目で把握できることが望ましい。ユーザは地図上で観光スポットを検索しているものとし、まず、地図上の実際にスポットが存在する座標に検索結果スポットを配置する。次に、既訪問スポットの位置を決定する。これは検索結果スポットとの対応関係によって変化し、検索結果スポットとの類似度が高いと近く、類似度が低いと遠くなるように配置することで、意味的な位置関係を表現する。既訪問スポットの座標は、式4.1によって求める。

$$Score(s_f, T) = \sum_{s_u \in S_u} w(s_u, s_f) \times P_{s_u, s_f}(d(s_u, T)) \quad (4.1)$$

$S_u$  を未訪問エリア内のスポット集合とする。 $w(s_u, s_f)$  は、ある既訪問スポット  $s_f$  と検索結果スポット  $s_u$  の類似度が正であれば 1.0、負であれば -0.5 を返す関数である。 $T$  は、候補の座標である。 $d(s_u, T)$  は、ある座標  $T$  から検索結果スポット  $s_u$  のユークリッド距離である。また、 $P_{s_u, s_f}$  は、平均  $\mu$  が 0、標準偏差  $\sigma$  が  $(1 - |\cos(\text{vec}(s_u), \text{vec}(s_f))|) \times \alpha$  である正規分布を表す。 $\alpha$  は重みである。 $\text{vec}$  は検索結果スポットや既訪問スポットの特徴ベクトルを返す関数である。 $Score(s_f, T)$  がもっとも大きな座標  $T$  を既訪問スポットの座標とする。

図4.1は既訪問スポットの経度座標計算の例である。縦軸はスコアの値であり、横軸は経度である。4つの検索結果スポットを用いて、 $Score(s_f, T)$ （合計スコア）を計算している。その結果、135.06の座標がもっとも高くなる。緯度についても同様に行う。

## 4.3 対応関係の可視化評価

### 4.3.1 実験内容

検索結果スポットの説明性向上のための観光スポットの対応関係可視化手法の評価を行うため、以下の3つの提示手法を使って比較する。図4.2、図4.3、図4.4は3つの提示手法それぞれのプロトタイプシステムとなっている。黒いマーカーは入力した既訪問スポット、赤いマーカーは未訪問エリア内のスポットを意味している。表示が重なることもあるため、黒いマーカーをクリックすると既訪問スポット名を吹き出しでも表示する。また、3つの提示手法において、類似度が類似度の平均以下であれば、関係を表現するキーワードが「なし」と

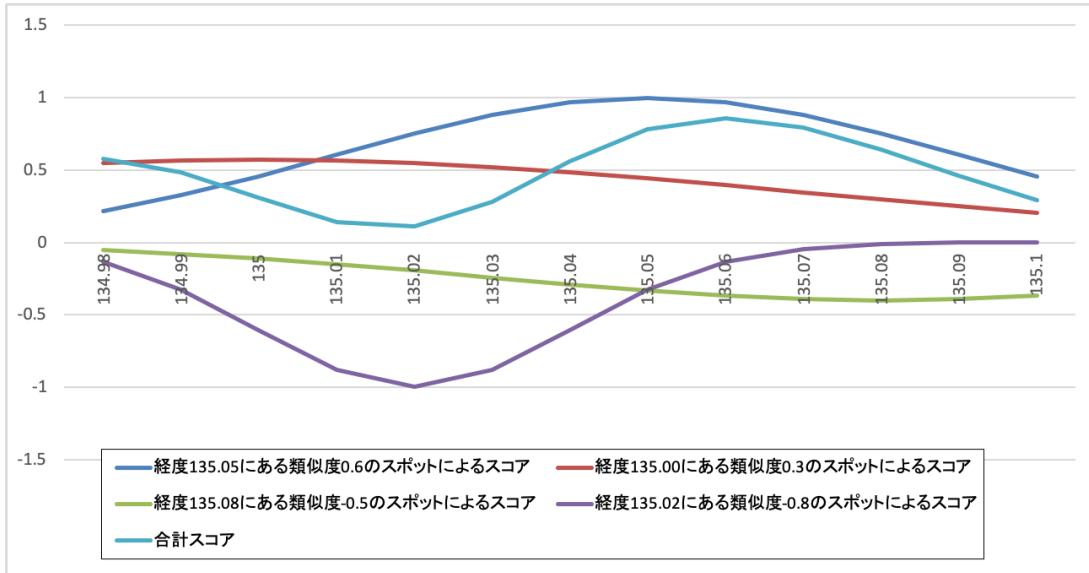


図 4.1 既訪問スポットの経度座標計算の例

表示する。これは意味が小さい関係を表すキーワードは一般的に理解しにくいためである。

*Position\_Map* (図 4.2) では、検索対象のスポットと既訪問スポットの関連度を全体的な位置関係で表現している。既訪問スポットは実際の座標ではなく、提案手法で決定した座標で表示される。赤いマーカーをクリックすると詳細情報が表示される。詳細情報は、検索結果スポット名、対応する既訪問スポット名、検索結果スポットと既訪問スポットの関係を表現するキーワードである。「既訪問スポット意味座標（表示/非表示）」ボタンは黒いマーカーの表示/非表示を切り替えることができる。

*Line\_Map* (図 4.3) では、位置関係による提示に加えて線の色を使って表現している。線の色は赤いマーカーと黒いマーカーの類似度を示している。線は緑色に近いと類似度が低く、赤色に近いと類似度が高いことを意味している。赤いマーカーをクリックすると詳細情報が表示される。詳細情報は、*Position\_Map*と同じく、検索結果スポット名、対応する既訪問スポット名、検索結果スポットと既訪問スポットの関係を表現するキーワードである。「既訪問スポット意味座標（表示/非表示）」ボタンは黒いマーカーの表示/非表示を切り替えることができる。

*Table\_Map* (図 4.4) では、位置関係や線は用いずに、*Line\_Map*で表現していたすべての情報を表形式でまとめて表現している。詳細情報は、現在選択している検索結果スポット名と、対応する既訪問スポット名、検索結果スポットと既訪問スポットの関係を表現するキーワード、類似度である。類似度の値は「-1.0 から 1.0」の間になる。「-1.0」はもっとも類似せず、「1.0」はもっとも類似することを意味している。

本実験では、近似として対象エリア内を緯度、経度それぞれ 17 個に分割し、等間隔で代



図 4.2 位置関係 (Position\_Map)



図 4.3 位置関係 + 線 (Line\_Map)



図 4.4 表形式 (Table\_Map)

表点を 289 個作り、それを既訪問スポットの候補点  $T$  として用いた。対応関係の可視化を評価するために、観光スポットの相対的特徴ベクトルによる対応関係と分散表現ベクトルによる対応関係の 2 つの手法を使って比較を行う。また、2016 年 9 月末までのじゃらん<sup>\*1</sup>から得られたレビューデータを用いる。

### 4.3.2 実験手順

クラウドソーシングのサービスである、 CrowdWorks を利用して被験者を集めた。分散表現ベクトルを用いたシステムでは、42 人合計 60 件のデータを集めた。相対的特徴ベクトル

\*1 <https://www.jalan.net/kankou/>

表 4.1 行き先の決定しやすさの回答数割合（分散表現ベクトル）

決定しやすさ	1	2	3	4	5	平均
<i>Position_Map</i>	0.0%	18.3%	11.7%	60.0%	10.0%	3.62
<i>Line_Map</i>	1.7%	25.0%	11.7%	43.3%	18.3%	3.52
<i>Table_Map</i>	0.0%	20.0%	23.3%	40.0%	16.7%	3.53

を用いたシステムでは、44人合計60件のデータを集めた。まず、被験者は3個から9個の自分がすでに訪れたことのある、かつ気に入った観光スポットを入力した。次に、被験者は旅行などで行ったことがなく、これから訪れたい都道府県やエリアを入力した。

本実験では、各提示手法において、検索スポットが重複しないように振り分け、3つの提示手法を連続して利用した。それぞれの提示手法において、被験者は検索スポットから行き先を決定した。このとき、表示順に影響をなくすために、6パターンの表示順を作成し、被験者はランダムに1つの表示順で実験を行う。実験を行った後、以下の項目について評価する。

- 各提示手法の行き先の決定しやすさ
- *Position\_Map* と *Line\_Map* どちらの方が、訪問履歴との関係がわかりやすかったか
- *Line\_Map* と *Table\_Map* どちらの方が、訪問履歴との関係がわかりやすかったか
- 各提示手法の訪問先決定までにかかった体感時間
- 各提示手法の訪問先決定までにかかった労力

また、それぞれの手法について、自由記述で意見を書いてもらった。

#### 4.3.3 分散表現ベクトルを用いた結果と考察

表 4.1 は3つの提示手法それぞれの行き先の決定しやすさを5段階評価した回答数である。5はとても決定しやすい、1はとても決定しづらいを意味している。表 4.1 の平均値から *Position\_Map* がもっとも決定しやすいことがわかった。*Position\_Map* は *Line\_Map* より決定しやすい理由として、*Line\_Map* は位置情報提示に加えて、線を描くことによって、操作が煩雑と感じる被験者がいるためと考えられる。ただし、3つの提示手法について  $t$  検定を行った結果、それぞれ有意差があるとはいえないかった。

表 4.2 は *Position\_Map* と *Line\_Map* どちらの方が、訪問履歴との関係がわかりやすかったかの回答数の割合となっている。表 4.2 から *Position\_Map* がわかりやすいの割合が 45.0% に対して、*Line\_Map* が 35.0% のため、*Position\_Map* の方がわかりやすいといえる。表 4.1 でも、*Line\_Map* がもっとも決定しづらいとなっている。「ラインが煩雑で見難いと

表 4.2 *Position\_Map* と *Line\_Map* の訪問履歴との関係の回答数割合（分散表現ベクトル）

回答	回答数の割合
<i>Position_Map</i> がわかりやすかった	21.7%
<i>Position_Map</i> がややわかりやすかった	23.3%
どちらでもない	20.0%
<i>Line_Map</i> がややわかりやすかった	25.0%
<i>Line_Map</i> がわかりやすかった	10.0%

表 4.3 *Line\_Map* と *Table\_Map* の訪問履歴との関係の回答数割合（分散表現ベクトル）

回答	回答数の割合
<i>Line_Map</i> がわかりやすかった	8.3%
<i>Line_Map</i> がややわかりやすかった	25.0%
どちらでもない	28.3%
<i>Table_Map</i> がややわかりやすかった	30.0%
<i>Table_Map</i> がわかりやすかった	8.3%

表 4.4 訪問先決定までにかかった体感時間の回答数割合（分散表現ベクトル）

	1	2	3	4	5	平均
<i>Position_Map</i>	13.3%	48.3%	20.0%	18.3%	0.0%	2.43
<i>Line_Map</i>	20.0%	36.7%	21.7%	20.0%	1.7%	2.47
<i>Table_Map</i>	13.3%	36.7%	30.0%	20.0%	0.0%	2.57

表 4.5 訪問先決定までにかかった労力の回答数割合（分散表現ベクトル）

	1	2	3	4	5	平均
<i>Position_Map</i>	23.3%	38.3%	21.7%	10.0%	6.7%	2.38
<i>Line_Map</i>	21.7%	31.7%	31.7%	11.7%	3.3%	2.43
<i>Table_Map</i>	16.7%	46.7%	25.0%	10.0%	1.7%	2.33

感じる.」といった意見の被験者が多数あり、強く好まない被験者が一定数が存在するため、このような結果となったと考える。また、表 4.3 は *Line\_Map* と *Table\_Map* の訪問履歴との関係のわかりやすさの回答数の割合となっている。表 4.3 から *Line\_Map* がわかりやすいの割合が 33.3% に対して、*Table\_Map* が 38.3% のため、*Table\_Map* の方がわかりやすい。

表 4.4 は 3 つ提示手法それぞれの訪問先を決定するまでにかかった体感時間を 5 段階評価した回答数である。1 は訪問先決定するまでに短い時間で、5 は長い時間で決定していることを意味している。表 4.5 は 3 つ提示手法それぞれの訪問先決定するまでにかかった労力を 5 段階評価した回答数である。1 は訪問先決定するまで少ない労力で、5 は多い労力で決定していることを意味している。体感時間と労力の相関係数を計算すると 0.73 と強い正の相関があり、体感時間が短いと労力も少なく、体感時間が長いと労力も多くなる。訪問先決定ま

表 4.6 行き先の決定しやすさの回答数割合（相対的特徴ベクトル）

決定しやすさ	1	2	3	4	5	平均
<i>Position_Map</i>	0.0%	10.0%	33.3%	46.7%	10.0%	3.57
<i>Line_Map</i>	3.3%	11.7%	18.3%	60.0%	6.7%	3.55
<i>Table_Map</i>	0.0%	11.7%	21.7%	48.3%	18.3%	3.73

表 4.7 *Position\_Map* と *Line\_Map* の訪問履歴との関係の回答数割合（相対的特徴ベクトル）

回答	回答数の割合
<i>Position_Map</i> がわかりやすかった	16.7%
<i>Position_Map</i> がややわかりやすかった	25.0%
どちらでもない	26.7%
<i>Line_Map</i> がややわかりやすかった	15.0%
<i>Line_Map</i> がわかりやすかった	16.7%

でにかかった体感時間が短いと回答した数の割合は *Position\_Map* が 61.7%, *Line\_Map* が 56.7%, *Table\_Map* が 50.0% となっている。よって、*Position\_Map* がかかった体感時間がもっとも短く、*Table\_Map* がもっとも長いことがわかった。訪問先決定までにかかった労力が少ないと回答した数の割合は *Position\_Map* が 61.7%, *Line\_Map* が 53.3%, *Table\_Map* が 63.3% となっている。よって、*Table\_Map* がかかった労力がもっとも少ない、*Line\_Map* がもっとも多いことがわかった。以上から、*Position\_Map* は体感時間が短くかつ労力もなく訪問先を決定することができ、*Line\_Map* はもっとも被験者にとって手間がかかっていることがわかった。対応関係を位置情報で表すことによって、表形式の時間と労力の問題点について改善することができるといえる。*Position\_Map* は被験者に提示する文字の情報を少なくして、位置関係を使って既訪問スポットと検索結果スポット対応関係を見せるこによって、他の 2 つの提示手法と比べると理解容易化することができたと考えられる。

#### 4.3.4 相対的特徴ベクトルを用いた結果と考察

表 4.6 は 3 つの提示手法それぞれの行き先の決定しやすさを 5 段階評価した回答数である。5 はとても決定しやすい、1 はとても決定しづらいを意味している。表 4.6 の平均値から *Table\_Map* がもっとも決定しやすいことがわかった。*Table\_Map* は *Position\_Map* と *Line\_Map* より決定しやすい理由として、*Table\_Map* はより日常的に使っている地図と類似するため使いやすいと感じている被験者が多いと考えられる。また、他の 2 つの提示手法では、位置情報提示や線を提示することによって、日常的に使わない地図の表現を用いたたつめ分かりづらいと考えられる。ただし、3 つの提示手法について *t* 検定を行った結果、それぞれ有意差があるとはいえないかった。

表 4.8 *Line\_Map* と *Table\_Map* の訪問履歴との関係の回答数割合（相対的特徴ベクトル）

回答	回答数の割合
<i>Line_Map</i> がわかりやすかった	6.7%
<i>Line_Map</i> がややわかりやすかった	15.0%
どちらでもない	38.3%
<i>Table_Map</i> がややわかりやすかった	31.7%
<i>Table_Map</i> がわかりやすかった	8.3%

表 4.9 訪問先決定までにかかった体感時間の回答数割合（相対的特徴ベクトル）

	1	2	3	4	5	平均
<i>Position_Map</i>	15.0%	28.3%	33.3%	23.3%	0.0%	2.65
<i>Line_Map</i>	8.3%	45.0%	21.7%	23.3%	1.7%	2.65
<i>Table_Map</i>	11.7%	43.3%	25.0%	18.3%	1.7%	2.55

表 4.10 訪問先決定までにかかった労力の回答数割合（相対的特徴ベクトル）

	1	2	3	4	5	平均
<i>Position_Map</i>	25.0%	23.3%	30.0%	18.3%	3.3%	2.52
<i>Line_Map</i>	13.3%	40.0%	23.3%	21.7%	1.7%	2.58
<i>Table_Map</i>	16.7%	43.3%	26.7%	11.7%	1.7%	2.38

表 4.7 は *Position\_Map* と *Table\_Map* どちらの方が、訪問履歴との関係がわかりやすかったかの回答数の割合となっている。表 4.7 から *Position\_Map* がわかりやすいの割合が 41.7% に対して、*Line\_Map* が 31.7% のため、*Position\_Map* の方がわかりやすいといえる。表 4.8 は *Line\_Map* と *Table\_Map* の訪問履歴との関係のわかりやすさの回答数の割合となっている。表 4.8 から *Line\_Map* がわかりやすいの割合が 21.7% に対して、*Table\_Map* が 40.0% のため、*Table\_Map* の方がわかりやすい。表 4.6 でも、*Table\_Map* がもっとも決定しやすい。被験者から「マーカーが密集しているため見づらい」などの意見多数あり、強く好まない被験者が一定数が存在するため、このような結果となったと考える。また、「黒いマーカーによる赤いマーカーとの関連性がわかりやすかった」や「色で関係性が表されていて、わかりやすかった」などの意見も多数あり、好む被験者も一定数が存在する。位置関係を表現する際に、仮想的な地物の類似性を視覚的に表していることが明確にわかるインターフェースにする必要がある。

表 4.9 は 3 つ提示手法それぞれの訪問先を決定するまでにかかった体感時間を 5 段階評価した回答数である。1 は訪問先決定するまでに短い時間で、5 は長い時間で決定していることを意味している。表 4.10 は 3 つ提示手法それぞれの訪問先決定するまでにかかった労力を 5 段階評価した回答数である。1 は訪問先決定するまで少ない労力で、5 は多い労力

で決定していることを意味している。体感時間と労力の相関係数を計算すると 0.73 と強い正の相関があり、体感時間が短いと労力も少なく、体感時間が長いと労力も多くなる。表 4.9 と表 4.10 から *Table\_Map* がもっとも被験者にとって手間がかからないことがわかる。訪問先決定までにかかった体感時間が短いと回答した数の割合は *Position\_Map* が 43.3%，*Line\_Map* が 53.3%，*Table\_Map* が 55.0% となっている。よって、*Table\_Map* がかかった体感時間がもっとも短く、*Position\_Map* がもっとも長いことがわかった。訪問先決定までにかかった労力が少ないと回答した数の割合は *Position\_Map* が 48.3%，*Line\_Map* が 53.3%，*Table\_Map* が 60.0% となっている。よって、*Table\_Map* がかかった労力がもっとも少ない、*Position\_Map* がもっとも多いことがわかった。以上から、*Table\_Map* は体感時間が短くかつ労力も少なく訪問先を決定することができ、*Position\_Map* はもっとも被験者にとって手間がかかっていることがわかった。

#### 4.3.5 分散表現ベクトルと相対的特徴ベクトルの比較

分散表現ベクトルによる対応付けと相対的特徴ベクトルによる対応付けの訪問先決定までにかかった体感時間と労力について比較を行う。訪問先決定までにかかった体感時間を表す表 4.4 と表 4.9 を比較すると、分散表現ベクトルの方がユーザにとって体感時間が短かいことがわかった。訪問先決定までにかかった労力を表す 4.5 と表 4.10 を比較すると、分散表現ベクトルの方がユーザにとって労力が少ないことがわかった。よって、分散表現ベクトルによって既訪問スポットと検索結果スポットで対応付けした方が被験者にとって体感時間が短く、かつ労力も少なく訪問先を決定することができる。また、分散表現ベクトルによる対応付けと相対的特徴ベクトルによる対応付けの訪問先決定までにかかった体感時間と労力について *Position\_Map* と *Line\_Map* の平均値の変動が、*Table\_Map* の平均値の変動より大きいことがわかった。

相対的特徴ベクトルによる対応付けについて、3 章での実験において、一般的に明らかな対応関係の抽出を苦手である結果が出ていることからも、一目で把握させる用途の可視化の対応関係としては不向きであると考える。そのため、*Position\_Map* と *Line\_Map* は被験者にとってわかりづらく、体感時間と労力が増えたものと考えられる。

## 4.4 まとめ

観光スポットを効果的に理解するために、対応関係に基づく可視化手法を提案した。評価として分散表現ベクトルによる対応関係の可視化と相対的特徴ベクトルによる対応関係の可視化に関する実験を行った。検索結果スポットを簡単に理解するために、入力した既訪問

スポットがどのような関係にあるのか地図上で、一目で把握できることが望ましい。そのために、2つのベクトルによるそれぞれの評価実験に対して、検索対象のスポットと既訪問スポットの関連度を全体的な位置関係で表現している *Position\_Map*, 位置関係による提示に加えて線の色を使って表現している *Line\_Map*, 位置関係や線は用いずに *Table\_Map* で表現していたすべての情報を表形式で表現している *Table\_Map* の3つの提示手法を比較することで評価を行った。

特徴ベクトルによる対応関係の可視化に関して *Position\_Map* がユーザにとってもっとも理解しやすいことがわかった。相対的特徴ベクトルによる対応関係の可視化に関して *Table\_Map* がユーザにとってもっとも理解しやすい結果となった。2つの対応関係の可視化手法について理解しやすいインターフェースはそれぞれ異なることがわかった。特徴ベクトルを使って観光スポットの特徴抽出と対応付けを行うことによって、観光スポット間の違いを直感的に地図上に表示することができる。

相対的特徴ベクトルは既訪問スポットと検索結果スポットの特徴の差によって対応付けるため、地図上で観光スポットの違いを位置情報で表示するとユーザにとって理解しづらいことがわかった。よって、相対的特徴ベクトルで対応付けした既訪問スポットと検索結果スポットの対応関係を地図上で位置情報として表示するに相応しくない。既訪問スポットと検索結果スポットの特徴の差を地図上で表示したい場合は、類似度を数値で表した *Table\_Map* の方がユーザにとって理解しやすい。

提案手法によって作った提示手法の *Position\_Map* と *Line\_Map* は対応関係の類似度を地図上で距離と座標で表しているため、相対的特徴ベクトルより分散表現ベクトルで対応付けした方がユーザにとって一目で把握でき、観光スポットを効果的に理解することにつながると考えられる。しかし、*Line\_Map* の方は位置情報以外に、カラー線で増加したため、ユーザにとって情報量が増加する。また、一般的に使っている地図検索システムのインターフェースと異なることによって、操作しづらいと感じるユーザも多数いる。

## 第 5 章

### 結論

観光地検索システムは、求める観光スポット情報を発見するために観光スポットの理解しやすいことが重要である。本研究では、観光スポットを効果的に理解するために、観光スポット間の対応付けと地図上における対応関係の可視化の 2 つの観点から研究を行っている。ユーザが訪れたことのある観光スポットを使って、検索結果スポットとの対応付けを行い、スポットの特徴を定義し、その対応関係を地図上に表示して、観光地検索システムの説明性を向上する手法を提案した。

観光スポット間の対応付けについて、既訪問スポットと検索結果スポットの関係を抽出するために、カテゴリなどのメタデータ、レビューの文書特徴量、さらに観光スポット間の特徴の差（相対的特徴）を用いた場合を比較した。観光スポットの特徴を抽出するために、観光スポット検索サイトのスポットレビューを利用した。ある観光スポットに対して、すでに訪問した観光スポットと比較することによって、その観光スポットの特徴ベクトルを作成し、特徴ベクトル間の類似度に基づいて既訪問スポットと検索結果スポットとの対応付けを行った。次に、ユーザに観光地を説明するために、観光スポット間の特徴語を抽出し、観光スポット間の対応関係を地図上で可視化する手法を提案した。また、それぞれに対して評価実験を行い、提案した各手法の有効性を評価した。

#### 1. 観光スポット間の対応付け

観光スポットの対応付け手法について、対応付け、特徴語抽出の評価を行った。まず、対応付けに関して以下に得られた知見を述べる。メタデータは明らかな関係にあるスポット同士を関連付けることができるが対応付け可能な数が少ない。分散表現ベクトルは、関係性を説明するキーワードは明確であるが、意外性のあるスポットとの関連付けしづらい。相対的特徴ベクトルを用いることによって、カテゴリをこえて各スポットの特徴を求めるため、明らかな関係となる対応付けが減らり、意外性のある対応付けが増えた。また、相対的特徴ベクトルを利用すると、カテゴリが異なる場合において

て、分散表現ベクトルよりもキーワードの明確性を増加することができた。特徴語抽出に関して、調和平均が他の2つの手法と比べると隠れた関係を表現する特徴語を見つけることが可能であり、無関係な特徴語を抽出しにくくなつたことから有効であると考えられる。また、検索結果スポット、対応する既訪問スポットと説明するための特徴語を提示することによって、ユーザが検索結果スポットに対する理解がしやすくなることがわかった。

## 2. 地図上における観光スポットの対応関係の可視化

観光スポットの対応関係可視化手法について、分散表現ベクトルと相対的特徴ベクトルによって対応付けされた既訪問スポットと検索結果スポットをそれぞれ3つの提示手法で評価を行った。分散表現ベクトルによる対応付けでは観光スポットごとの特徴の類似度をはかり地図上に表示することで、提案手法である *Position\_Map* がユーザにとって一目で観光スポット間の関係性を把握することができ、観光スポットを効果的に理解することにつながる。相対的特徴ベクトルによる対応付けでは観光スポット間の特徴の差の類似度をはかり地図上に表示することで、提案手法ではうまく観光スポット間の関係性を表示できないため、*Table\_Map* のように類似度を数値で表した方がユーザにとって理解しやすいことがわかった。

以上のことより、観光スポットの対応付けに関して、相対的特徴ベクトルを利用して抽出された特徴語をユーザに提示することによって、既訪問スポットと検索結果スポットの関係をより詳細に理解することができる。しかし、相対的特徴ベクトルで対応付けした既訪問スポットと検索結果スポットの対応関係を地図上で可視化する際に、観光スポット間の差の類似度を表現しているため、ユーザにとって理解しづらい問題がある。そのため、地図上で可視化する際に、分散表現ベクトルにより対応付けした方が相対的特徴ベクトルで対応付けした既訪問スポットと検索結果スポットの対応関係より理解しやすいことがわかった。

今後の課題として、観光地検索システムにおいて、検索結果スポットに対する説明の理解を向上することができたが、ユーザの利用場面に応じた改良案について検討することができると考えられる。さらに、地図アプリや乗換案内といった Web ページ以外の異種メディアにおける対応付けに関して検討する必要がある。ユーザが情報を収集する際、地図アプリによる位置関係やルートの確認、乗換案内アプリによる地域移動時における交通情報の収集などの行動が想定される。これらの分野についても研究を進めることで、より観光地検索システムがユーザにとって有用なシステムになると考えられる。

## 謝辞

修士課程での研究全般に渡ってご指導賜りました北山大輔准教授に厚く御礼申し上げます。

お忙しい中貴重な時間を割いて副指導教員および副査を引き受けてくださり、ご指導賜りました三木良雄教授に感謝致します。

同様に、お忙しい中貴重な時間を割いて副査を引き受けてくださり、ご指導賜りました橋完太准教授に感謝致します。

研究室配属からの3年間に渡って、切磋琢磨し心の支えにもなってくれた歴代のインタラクティブメディア研究室メンバ全員に感謝致します。

最後に、名前を書き切ることはできませんが、両親をはじめ、私のこれまでの人生を支えてくださった全ての方に心から感謝いたします。ありがとうございました。

# 本研究に関する発表

## 論文

- [1] 潘健太, 北山大輔, “説明性向上のためのユーザレビューを用いた観光スポットの対応付け手法,” 情報処理学会論文誌：データベース, 第 84 号, 2020. (採録決定)

## 査読付き会議

- [2] Kenta Han, Daisuke Kitayama, “An Explanation Method of Unfamiliar Tourist Spots based on Roles of User’s Familiar Spots,” Lecture Notes in Engineering and Computer Science: Proceedings of The International MultiConference of Engineers and Computer Scientists 2019, pp. 358-362, March, 2019.

## 研究会

- [3] 潘健太, 北山大輔, “ユーザの既訪問スポットの位置付けに基づく検索結果スポットの説明手法,” 第 11 回データ工学と情報マネジメントに関するフォーラム (DEIM Forum 2019), H7-2, 2019
- [4] 潘健太, 北山大輔, “地図上における検索結果スポットの説明性向上のための観光スポットの対応関係可視化手法,” 観光情報学会第 20 回研究発表会論文集, pp. 5-8, 2019

# その他の発表実績

## 研究会

- [1] 潘健太, 北山大輔, “ユーザのレビュー選択に基づく観光スポット検索手法,” 第 10 回 データ工学と情報マネジメントに関するフォーラム (DEIM Forum 2018), H1-2, 2018
- [2] 潘健太, 北山大輔, “観光スポット検索のためのユーザのレビュー選択と特徴抽出に関する考察,” 電子情報通信学会技術研究報告, Vol. 118, No. 107, pp. 21-24, 2018

# 受賞

## 査読付き会議

- [1] Kenta Han, Daisuke Kitayama, “An Explanation Method of Unfamiliar Tourist Spots based on Roles of User’s Familiar Spots,” Lecture Notes in Engineering and Computer Science: Proceedings of The International MultiConference of Engineers and Computer Scientists 2019, pp358-362, March, 2019.